

Healthcare Twitter Analytics

Have a look at the relationships among the numerical fields

```
file = 'Tweets_Celiac_sent.csv'
data = read.csv(file)
data=data[c("id","retweet_count","listed_count","favourites_count","sentiment","score")]
first.look(data[c(-1)])
```

```
## ----- str: look at data types and values -----
## 'data.frame':    5675 obs. of  5 variables:
##  $ retweet_count   : int  10 0 0 0 2 1 0 0 2 4 ...
##  $ listed_count    : int  142 5 5 5 27 71 71 13 59 223 ...
##  $ favourites_count: int   55 0 0 0 93 0 0 482 62 55 ...
##  $ sentiment       : num   0 0 5 0 3 2 3 3 3 0 ...
##  $ score            : num   7.73 6.84 6.88 6.86 8.4 ...
## NULL
##
## ----- sample size: 5675
## ----- sample size of complete cases: 5675
## ----- difference: 0
##
## ----- names: variable names
## [1] "retweet_count"      "listed_count"       "favourites_count"
## [4] "sentiment"          "score"
##
## ----- some: Randomly select a few elements
##      retweet_count listed_count favourites_count sentiment  score
## 330                0            0             965          7  7.427
## 446                0            5             204          2  7.082
## 1088               3          1547          27823          4 11.687
## 3164               1           123           2187          4 11.289
## 3610               0            2             359          7  7.038
## 3815               3           499             605          7 11.913
## 4517               0            87          1485          3  7.159
## 5019               6            4             464         -3  7.434
## 5100               0            1              0          8  7.123
## 5110               0            1              0          8  7.130
##
## ----- summary: statistics for each variable
##  retweet_count  listed_count  favourites_count  sentiment
##  Min.   : 0.00   Min.   :    0   Min.   :    0   Min.   : -9.00
##  1st Qu.: 0.00   1st Qu.:    4   1st Qu.:   36   1st Qu.:  0.00
```

```

## Median : 0.00 Median : 16 Median : 228 Median : 0.00
## Mean : 1.78 Mean : 80 Mean : 1377 Mean : 1.21
## 3rd Qu.: 2.00 3rd Qu.: 59 3rd Qu.: 998 3rd Qu.: 2.00
## Max. :40.00 Max. :30761 Max. :108340 Max. :16.00
## score
## Min. : 6.22
## 1st Qu.: 6.97
## Median : 7.33
## Mean : 8.02
## 3rd Qu.: 8.50
## Max. :13.91
##
## ----- grfsummary: statistics for each quantitative variable
## retweet_count listed_count favourites_count sentiment score
## n 5675.00 5675.00 5675.00 5675.00 5675.00
## nNA 0.00 0.00 0.00 0.00 0.00
## min 0.00 0.00 0.00 -9.00 6.22
## mean 1.78 79.68 1376.79 1.21 8.02
## median 0.00 16.00 228.00 0.00 7.33
## stdev 4.09 627.11 5633.75 2.34 1.55
## skew 4.65 42.81 13.79 0.86 1.33
## npskew 0.43 0.10 0.20 0.52 0.44
## kurtosis 27.33 2036.84 240.79 2.05 0.61
## max 40.00 30761.00 108340.00 16.00 13.92
##
##
## ----- possible factor non-independence -----
## factor variable names:
## List of possible factor pairs with interactions:

```

